Curator's Workbench

A Desktop Tool for Preparing Digital Materials
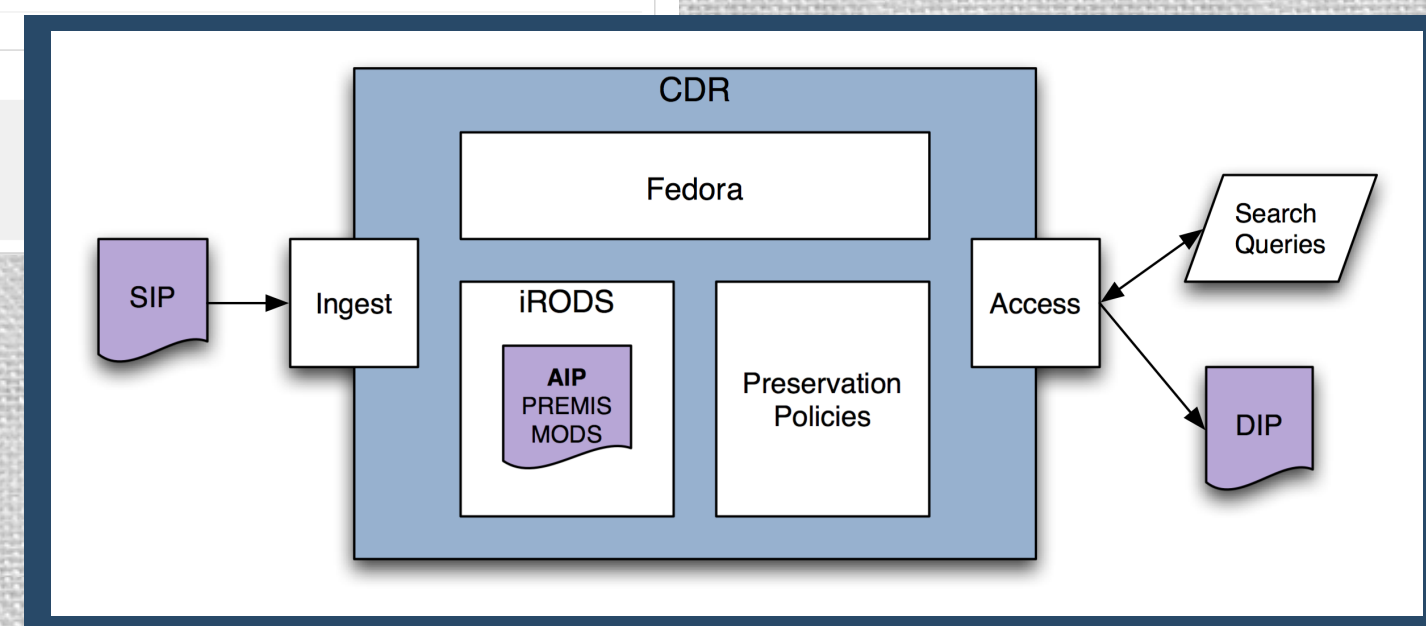
Erin O'Meara
Gregory Neil Jansen

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

CAROLINADIGITALREPOSITORY

Overview

The Curator's Workbench is a new collection preparation and work flow tool for digital materials, developed in summer 2010 at UNC Chapel Hill. It is designed to streamline the capture, arrangement and description of digital materials for an institutional repository, while automating routine steps, such as staging files and calculating checksums.

About the Project

The Curator's Workbench software was created for the Carolina Digital Repository, an initiative of the University Libraries of UNC Chapel Hill.

The workbench continues to be developed to support the work flow needs of UNC Libraries. However the software is available for community experimentation and use. The software has a modular design and may be extended for use outside of the UNC environment.

Contact Us:

METS Manifest

Each project in the workbench is built around an underlying METS manifest. This manifest tracks all selected digital objects, their replicas and their metadata. When the project is ready for submission, an export function translates the internal METS into a submission package ready for ingest.

Screenshot 1: Editing a Description

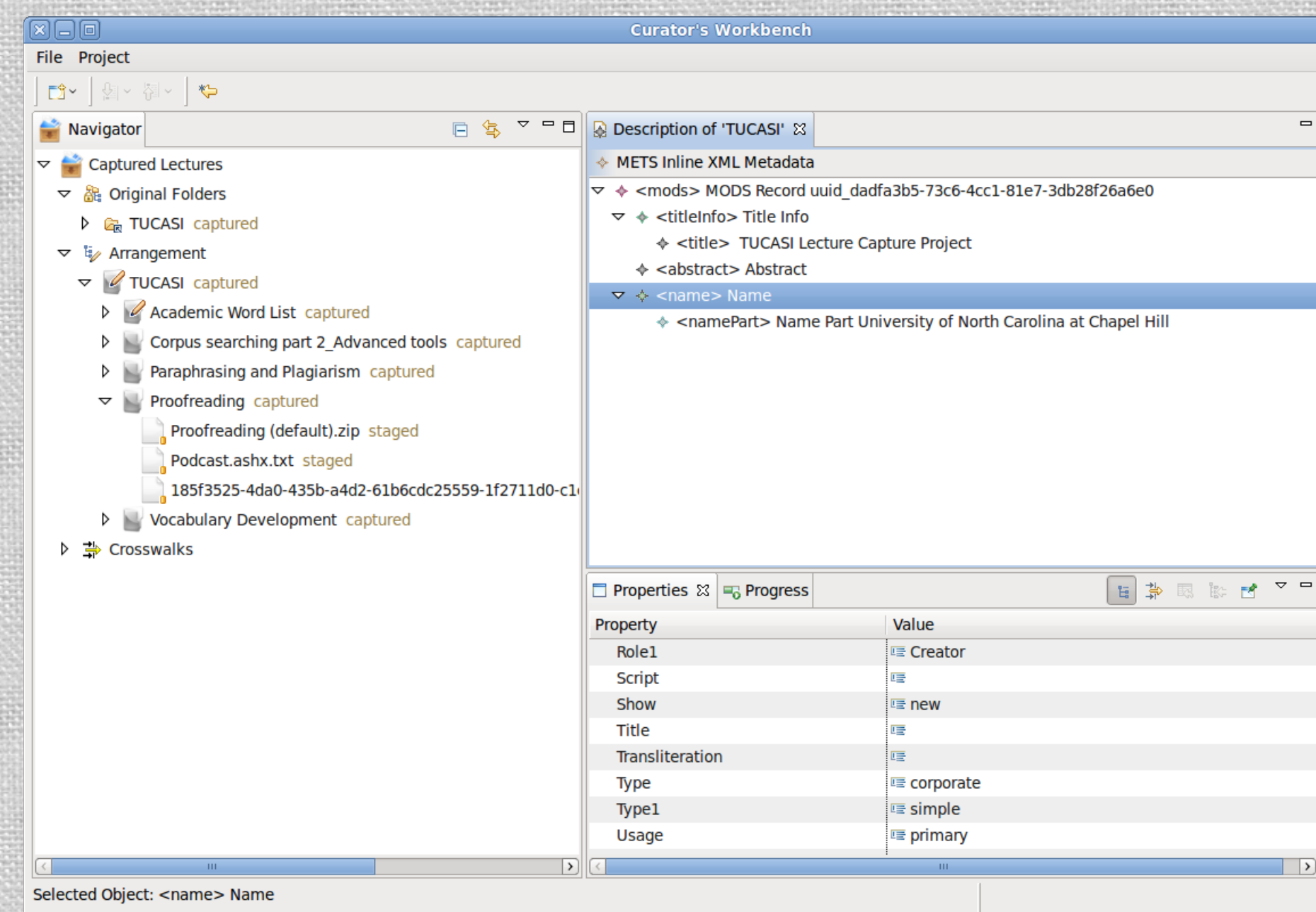### Screenshot 2: Defining a Crosswalk

File Staging

As the user selects and rearranges files, the workbench works in the background to "paranoid" copy the selected files to a staging location. This advance staging makes it possible to process large SIPs without having to unreliably transport data files at the time of submission.
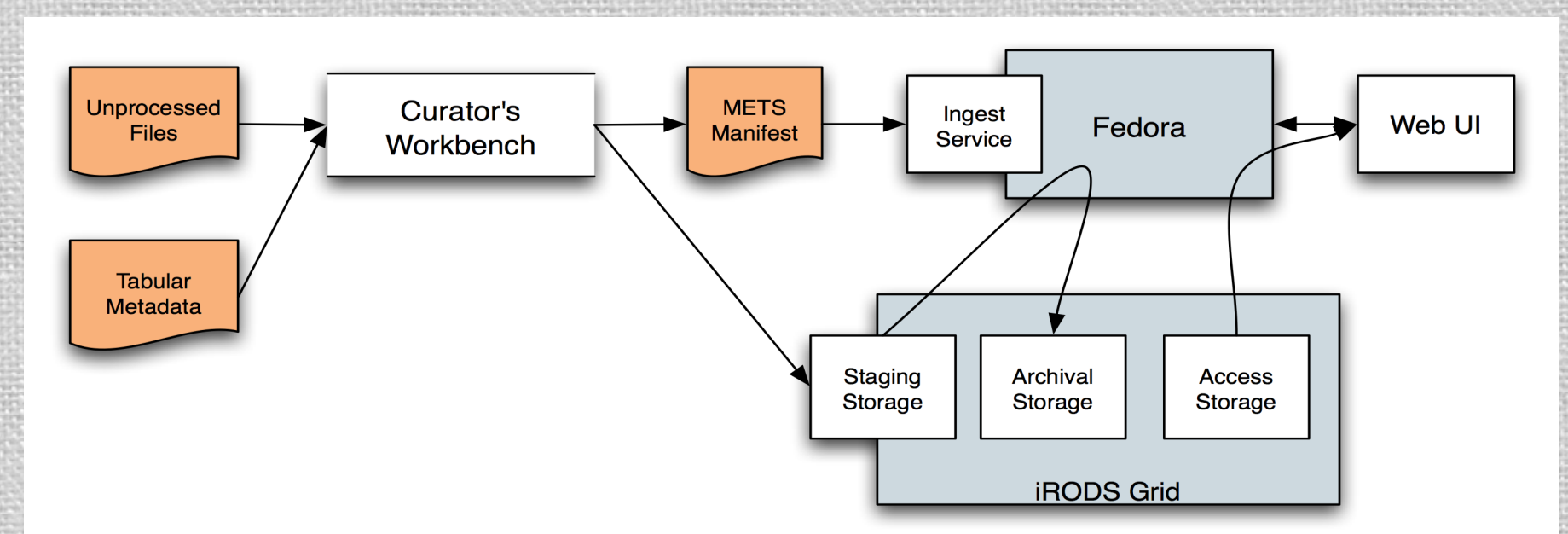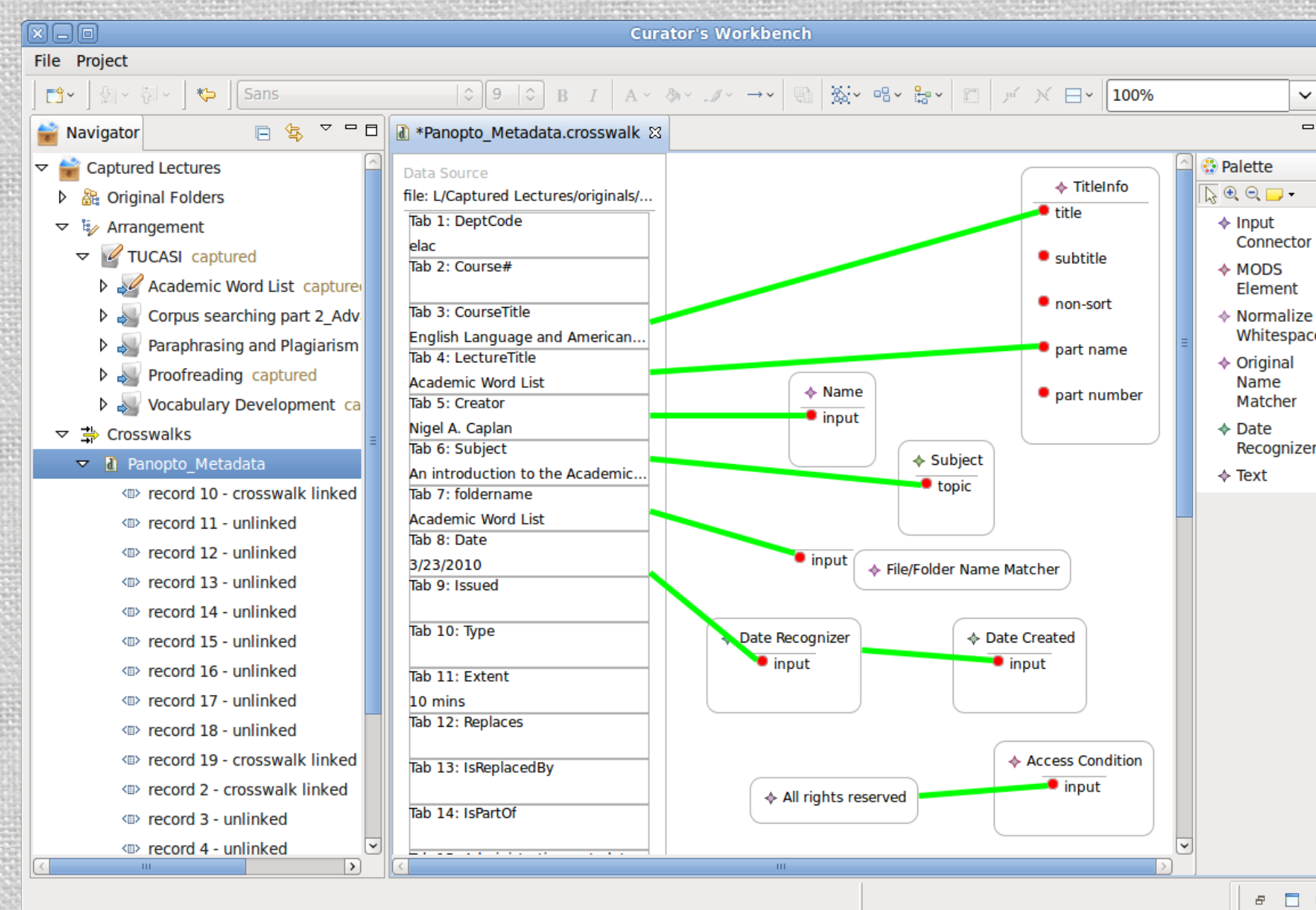
The staging location is configurable. In the case of the CDR, the staging location is within our iRODS data grid, as shown in the figure below. Staged files are checksummed and replicated to a disk resource, but not copied to tape until ingest.

The workbench starts staging as soon as the first file is selected for capture. The files have indicators to show which are staged and which are still queued for staging. Users can also monitor the progress of staging operations in a progress view.

Workbench in Submission Work Flow

Metadata Crosswalks

We built this tool to remove a key bottleneck in ingest work flow, which was the writing of custom scripts to migrate user supplied metadata to our standard schema, MODS. This cumbersome and arcane scripting process was replaced with a visual programming model, which allows non-programmers to define how incoming tabular fields map to MODS elements.

An example crosswalk definition is shown in the screenshot to the left. Green lines connect the tabular data fields on the left with MODS elements on the canvas. Sometimes an intermediate operation is defined, such as the "Date Recognizer". Standard text like "All rights reserved" can also be defined and mapped, to be inserted into every MODS record.

A special widget, called a "File or Folder Name Matcher" is used in this example to automatically match crosswalked records to files selected in the workbench. Records that are still unmatched can be linked by drag and drop.

The crosswalked records and automatic links are regenerated every time a crosswalk definition is modified.